DOCUMENT RESUME

ED 190 623 TM 800 417

AUTHOR Anderson, Scarvia B.
TITLE Going Public: A Message for the Measurement Community.
PUB DATE Mar 80
NOTE 16p.: Paper presented at the Conference on Measurement and Evaluation (Los Angeles, CA, 1980) and the Annual Meeting of the Association for Measurement and Evaluation in Guidance (Atlanta, GA, March 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Accountability; Change Strategies; *Communication Skills; Court Litigation; Credibility; Disclosure; *Evaluators; Legal Responsibility; *Public Policy; *Standardized Tests; State Legislation
IDENTIFIERS Test Publishers; Truth in Testing Legislation

ABSTRACT

Measurement and evaluation have become major public issues in the legislatures and the courts. Testing legislation represents two extremes: (1) confidence in external objective measurement for high school teacher and occupational certification; and (2) distrust of the results. Since 1978, truth-in-testing bills for admissions or other standardized tests have been proposed or enacted in several states; most resemble the Neville bill enacted in New York state. Measurement specialists have been summoned to testify in the courts, as well as the legislatures, concerning the validity of employment and educational tests. Professionals must accept their new public role, by learning to communicate with the public. They should learn to discriminate between sincere interest in understanding and improving testing, versus advocacy of a position that all tests are biased and wasteful. Secondly, they should simplify communication by explaning only a few measurement concepts, such as aptitude, achievement, error of measurement, norms, correlation, and restriction of range. Finally, because test results are often vaguely stated, specialists should either give the best answer or admit they don't know or have no comment. (CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
S. Anderson
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED190623

GOING PUBLIC

A Message for the Measurement Community*

Scarvia B. Anderson
Educational Testing Service

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Most of us in the measurement community got our training at a time when tests were accepted as a necessary inconvenience (at worst), and the competencies and motivations of testers and evaluators were rarely questioned.

A few of us even entered the profession when:

- IQ tests were used in almost all elementary schools, although pupils' scores were not revealed to their parents;
- College Board scores were reported to colleges but not to students;
- the Coleman report, which used tests as a measure of the quality of schooling, had not been published;
- the massive educational intervention programs that would rely for their evaluation at least partly on tests had not been mounted; and
- there was no educational accountability movement -- the word 'accountability' had not even crossed fields.

We were hardly naive about the problems and claims of testing, but it was a relatively new science and we were challenged by that very fact. Not only were we challenged; most of us also were inspired by a sense of contributing to society -- of bringing order out of chaos, fairness out of privilege, objective truth out of subjective judgments.

TM 800417

---

*From speeches at the Conference on Measurement and Evaluation, Los Angeles, and the meeting of the Association for Measurement and Evaluation in Guidance (AMEG), Atlanta, March 1980.

During those halcyon days, we learned about KR 20 reliabilities, equating and scaling methods, and how to assemble an exam of appropriate difficulty. We subscribed vigorously to the APA standards for educational and psychological tests, debated fine points of sampling with out colleagues, and foresaw a future in which we would continue on these same paths, serving educational and other social institutions through measurement. We would help extend freedom of choice and opportunity to diverse and unprecedented numbers of recipients, and we would carry out our mission efficiently, professionally, and quietly.

But suddenly a new era was upon us. The accuracy of predictions about our professional life style plummeted. Measurement and evaluation made the news, as the target of political forces and a rallying point for special interest groups.

We have a long history of explaining away any dissatisfactions with the educational "system" by blaming successively vulnerable parties: first we blamed the students, then we blamed their parents (indicting Dr. Spock along the way), and then we blamed the teachers. But even after the sixties, inequities still prevailed, apparently unresponsive -- at least immediately -- to either legislation or innovation. It was clearly time to nominate a new villain, and no group was more anxious to serve on the search committee than those who presently occupied the chair -- the teachers.

The SAT score decline provided reinforcement for the back-to-basics movement; and, although the College Board and ETS protested that SAT scores were not designed to evaluate schools and the Wirtz "blue-ribbon" panel found complex and interwoven causes for the score drops, many groups hit upon the SAT decline as just one more sign that the schools and teachers were not doing their jobs. Minimum competency testing programs began to pop up in state after state.

As early as 1972, the leadership of the National Educational Association called for a moratorium on the use of standardized tests. And, speaking last fall to an assembly convened to devise strategies for "testing reform," NEA director Terry Herndon called for national legislation to regulate testing and testers. For the record, the other major organized teachers' group, the American Federation of Teachers, favors testing and opposes anti-testing legislation. There is also considerable doubt that rank-and-file NEA members are as strongly anti-testing as their national officials. However, the national NEA is not the only group that would do away with most -- if not all -- testing.

It is a fact that, on the average, black students score significantly lower than white students on nationally standardized tests. In a recent five-year period, black students "who took the SAT scored an average of 119 points lower than whites on the verbal section and 134 points lower than whites on the mathematics section" (Jacobson, 1980, p. 5). Back in 1969, the Association of Black Psychologists, charging cultural bias, called "for a moratorium on the repeated abuse and misuse of the so-called conventional psychological tests" (cited in Williams, 1974, p. 17). And at its 1975 convention, the NAACP resolved to "demand a moratorium on standardized testing whenever such tests have not been corrected for cultural bias" and recommended that "administrative and legal remedies" be employed to prevent test misuse (cited in the NAACP report on minority testing, 1976, p. 2).

Among the most prominent opponents of testing are Ralph Nader and his campus-based Public Interest Research Groups. They allude to inappropriate applications of tests, but aim their biggest guns at so-called monopolistic testing organizations -- one in particular.

Natural concomitants of all of this ferment have been legislation and court cases involving tests. When individuals and institutions cannot seem to solve their problems, there is an inevitable tendency to turn to government. As might be expected, the legislative initiatives involving testing have represented two extremes: on the one hand, confidence in external, objective measurement and, on the other, distrust of the results. Thus, we find increasing attempts to grant or deny credentials on the basis of mandated examinations: for high school students to receive "proper" diplomas, for undergraduates to avoid "remedial" classes, for teachers to renew or retain their status, and for persons wanting to enter or stay in a wide variety of other professions. Statewide legislation of this type has been enacted, for example, in Florida. Their 1976 Educational Accountability Act places a ban on "social promotion" and mandates the use of standards and tests for high school graduation, grade promotion, and early exit from high school.

Similarly, those authorizing large expenditures for educational, vocational, and other social innovations at national and state levels are increasingly explicit in their laws and regulations about the need for periodic appraisal -- testing, if you will -- of the effects of those programs, to determine whether they are reaching and serving the needs of the populations they were designed for. Regulations pertaining to Title I of the Elementary and Secondary Education Act offer some of the most clear-cut and prescriptive requirements for such evaluation.

As I have indicated, other legislation seems motivated by distrust of testing. The first state legislation directed specifically toward educational admissions testing was enacted in California in 1978. It requires test producers to provide test takers with sample tests with

questions and answers "equivalent to those in use" and to report certain actuarial data pertaining to numbers of test takers and so on. In fact, the California regulations do not provide any greater service for test takers than was already being provided voluntarily by testing organizations; the regulations simply introduce an additional reporting line and provide an additional repository of information.

Since 1978, bills to regulate practices associated with the administration and use of standardized tests in education have been proposed or introduced in a number of states -- the exact number varies from day to day. Most are patterned after the LaValle bill in New York state. That law, though evidently intended to benefit those who take various kinds of admissions tests, seems destined to result in hardships for the very persons it was intended to help. These hardships include increased costs to all test takers (the result of increased costs of producing the testing materials required by the law) and inconveniences to many (especially the handicapped and persons who for religious reasons are unable to take tests at regular Saturday administrations). The law makes it prohibitive to provide these special groups with as many extra administrations as were possible in the past. The New York law also introduces serious questions of extraterritoriality (can one state enact laws that impinge on the rights of citizens of other states?), possible violation of the Federal Copyright Act, and perhaps deprivation of property without due process of law.

Of the bills proposed or introduced, the most extreme and encompassing one is the law proposed in New Jersey. That law, if passed, would cover all standardized tests developed for the purpose of selection, placement, classification, graduation or any other bona fide use with students in

elementary, secondary, postsecondary, and professional schools. Further, it would not only require the disclosure of all test forms and answers immediately after use, as in the New York law, but would also prohibit testing organizations from charging any extra fee for this service.

Two bills to regulate standardized testing were proposed at the national level last year. The first, introduced by Congressman Weiss of New York, closely paralleled the LaValle bill. Congressman Weiss, with support from Nader forces, the NEA, and others, has promised to bring it back for reconsideration this year. The other bill, introduced by Congressman Gibbons of Florida, was especially notable in its attempt to legislate appropriate psychometric practice; for example, under such-and-such circumstances you must use a criterion-referenced test.

Concerned spectators at these legislative games are confronted with a peculiar dilemma. If legislation becomes a foregone conclusion, could we as professionals relate better to a single set of federal regulations than to a multiplicity of conflicting state regulations? But in the United States we have traditionally cherished the rights of states and individual institutions to set their own standards and make their own decisions in educational policies and practices. Any federal regulation would inevitably impinge on these rights.

Meanwhile, measurement specialists have been trooping to the courthouses as well as the halls of legislation. Five of the most famous cases involving employment tests are probably Griggs v. Duke Power, Davis v. Washington, U.S. v. State of North Carolina, U.S. v. State of South Carolina, and U.S. v. Detroit Edison. The first three pertained to the job-relatedness of selection measures; the North Carolina case dealt additionally with the use of cutoff scores; and the Detroit Edison case focused on test security and

confidentiality of individual results. Most professionals agree that in all five cases the court reached the "right" decision in terms of both measurement and human values. However, in a brilliant paper on the "War on Testing," psychologist and lawyer Barbara Lerner (1979) notes with respect to the Detroit Edison case:

> Let there be no confusion about my point here: I applaud the result the Court reached; what I deplore are the reasons the Court gave for reaching it because they set no limit on the personal predilections of individual judges and provide no firm basis for predicting what the Court will do next time (p. 6).

Two of the most famous court cases involving measurement in education came from California and are known familiarly as the "Bakke" and "Larry P." cases. In the former, the U.S. Supreme Court produced one of its more confusing sets of judgments, both ordering the admission to medical school of the white Mr. Bakke, who charged that he had been discriminated against, and upholding the use of race as a factor in university admissions programs. In the "Larry P." case, U.S. Judge Peckham concluded, against the wisdom of the measurement profession, that a difference in scores between two groups must prima facie indicate discrimination. However, in a third case, the "Debra P." case challenging the Florida Educational Accountability Act, the judge's ruling was that it was the educational system that was biased, not the functional literacy test in question.

The accumulated evidence is plain. There are no longer any cloisters or even closets where professionals in measurement and evaluation can quietly pursue their work. Measurement and evaluation

have become major public issues. The practice of measurement and evaluation is perforce a public activity. What does this mean for our daily work and the survival of our profession?

First, it means that we must accept our public roles, as an important part of our professional responsibilities.

Second, we must learn how to play those public roles effectively.

Third, in addition to teaching our graduate students and interns the principles and techniques of measurement and evaluation, we must initiate them into their public roles.

Fourth, we must not be deflected from simultaneously and continuously striving to improve our methods for assessing individuals and social programs.

Acceptance of our public roles is no mean acomplishment in itself, for it admits first of all that we and our profession, and the institutions we represent, are not universally understood, respected, or loved.

Learning how to play our public roles is even harder. The academic rules by which we were brought up are not relevant. We were taught to respect reason and to show respect even for our colleagues whose reasoning we found faulty. Our controversies were aired at professional meetings in a private language that we and our mentors invented. If another professional published a journal piece that we considered scientifically outrageous, we could get a rebuttal in an ensuing issue that went to the same readership. Our controversies seldom impugned the motives of our opponents. We learned to accumulate evidence to support our conclusions and were frequently content to take a long time doing it.

But hardly any of these strategies are effective in dealing with an angry parent whose high school son's reading scores are abysmal; or with a teacher who is scared to death that he will be penalized for his class's low math performance; or with a student suspected of cheating on a test; or a newspaper reporter who wants to rank the quality of schools in the city on the basis of average SAT scores; or a judge who is convinced that employment tests are racially biased; or an agency with a large investment in a new training program, confronted with evidence that the program isn't nearly as effective as advertised; or a consumer advocacy group that sees tests as as good a cause as any to stimulate a flagging movement; or a legislator who sees a new and newsworthy opportunity in issues related to educational testing.

The angry parent is not going to sit still for an explanation of the differences between percentiles and percentages. The reporter wants a story, uncomplicated by questions of differing characteristics of students when they entered the schols or unrepresentativeness of the sample of students who self-selected to take the SAT. The judge who has had long experience in dealing with prejudice is going to have difficulty understanding that differences in performance between groups do not necessarily imply bias. And the legislator: A lawyer acquaintance noted recently that the average attention span of a legislator was shorter than that of a fruit fly in heat. Not a very flattering metaphor and not true of many legislators, I am sure. But it is true that their agendas are far different from, and usually much more complicated than, those of the measurement specialist.

Communicating with anyone outside the field is especially troublesome when what we have to say seems counter to common sense or personal experience. For example, the so-called "truth-in-testing" movement

advocates disclosing all standardized test forms after they are given. How can you explain to students -- or anyone -- why they shouldn't get their corrected test papers back? They got them back in school. Processing problems when thousands of students are involved, difficulties in test development and scaling, an increased emphasis on testing rather than the hoped-for deemphasis, even increased costs all seem very abstract. Another example: (This is the kind of thing that used to happen to me when I was involved in national norming administrations.) An angry assistant superintendent calls up and says "What do you mean these are national norms! The only school you selected in Cleveland was Jones School. Now if you had selected..."

Trying to have a sane discussion about coaching for tests runs into the same kinds of problems:

The Measurement Specialist says: "The effect of coaching on the Scholastic Aptitude Test has typically been a gain of less than 15 points for the verbal section and of less than 20 points for the mathematics section on the SAT scale, which ranges from 200 to 800."

The Layman says: "Are you trying to tell me that students shouldn't study for these tests? That doesn't make any sense. I always studied for tests."

Measurement Specialist: "I was talking about coaching not studying..."

Layman: "What?"

Measurement Specialist: "Well, you see if a student gets a book and maybe a teacher and really learns more about fractions and decimals, that's studying, but if he just drills on the kind of items on the tests, that's..."

Layman: "How can you tell which is which?"

Measurement Specialist: "It depends, of course, on how intensive or extensive the training is and whether the student is really learning the content or... Besides, the situation is somewhat different for achievement tests."

Layman: "What? All I want to know is whether I should send Mary Jane to a coaching school."

Measurement Specialist: "I am not in a position to make any recommendation about Mary Jane. I don't know, for example, what her abilities and motivation are. I don't know about the coaching course you're considering. What I can tell you is that on the average..."

Layman: "Average, smaverage..."

And so it goes.

The opponents of scientific measurement and evaluation have been quite effective in reducing their positions to slogans. Let me take a lesson from their books and offer in slogan form three pieces of advice about interacting with the media, legislators, citizens, and other groups on measurement and evaluation issues:

1. Recognize motivation.
2. Simplify communication.
3. Minimize equivocation.

By recognizing motivation, I mean trying to discern what the other party is really interested in or what self-interest he or she is trying to promote. Some individuals and groups are sincerely interested in understanding or improving a measurement process or application. The codes that should guide our behavior with them are similar to those we use with our professional colleagues -- except, of course, we need to remember the other two pieces of advice, "Simplify communication" and

"Reduce equivocation." With these groups, as with out colleagues, we must also be constantly mindful of the inter in interaction. They can teach us things and focus our attention on concerns that they, as consumers, have. We must listen.

Other individuals or groups are frankly advocates for a position -- that all assessment techniques are biased, that college admissions practices should be regulated by the government and not left to institutions, that money should not be invested in evaluating a particular program, that the schools in this district are inadequate, that tests should be abolished. We lose every time if we fail to discern their true motivation and respond to that, instead of to a cloak of rhetoric or to a question we hope they're concerned about because we know the answer. We're going to lose a lot of skirmishes anyway, of course, for it is going to take us a long time to adapt to the style of political advocacy where the meanings of "research," "study," "conclusion," and other words are so different from their meanings in the scientific context with which we are familiar and comfortable.

The first step in simplifying our communication is not to try to tell the public everything we know. In measurement, for example, we could create a far better climate for appropriate practice and use if we could get people to understand a very small number of concepts. My list would probably start with aptitude, achievement, error of measurement, norms, correlation, and restriction of range. Then we must consider how to explain these concepts, the language and comparisons that will be meaningful in light of the audience's experience and preconceptions. For example, the concepts of "correlation" and "restriction of range" can be made meaningful to some audiences by examining the relationship between

height and ability-to-score-baskets for a general school population and for players in the NBA. And in many cases we would be ill-advised to launch into a discussion of errors of measurement in test scores without first exploring the audience's appreciation of errors in other kinds of measurement; some people believe that only the measures we develop have this property!

The dialogue about coaching was full of the kind of equivocation that is so frequent in our communications that we're unaware of it. "It would seem that," "on the other hand," "depending on which definition you adopt," "there is a suggestion that," and, of course, "further research is needed" are firmly entrenched in our professional vocabularies and even in our manners. This is certainly appropriate in a scientific environment, but it is confusing and suspect to nonprofessional groups. People expect to get answers from specialists. They may not be satisfied with the answers they get and thus choose to seek "another opinion," but they want answers. Think about your own expectations from physicians, engineers, tax accountants, and automechanics. We are simply going to have to grit our teeth and give the best answer we know, or else say we don't know or "no comment." All three of these can get us into trouble: the "best" answer if it later proves wrong in a particular case, "don't know" because it reduces confidence in us, and "no comment" because it looks as if we're hiding something. But those risks seem less serious in the long run than the risk we run every day of not communicating at all if we stick to our old equivocal style when dealing with the public and public issues.

One of the most potent weapons we have in dealing with out-and-out opposition to scientific and objective measurement and evaluation is a challenge to consider the social consequences of not measuring or

evaluating. Of those who would do away with tests -- for example in college admissions -- we must ask what they would substitute. Interviews, character references, what prep schools the candidates attended? Would they rely exclusively on inflated and frequently biased high school grades? Of those who would dismiss objective evaluations of the effectiveness of social and educational programs, we must ask what alternatives they would propose to find out if taxpayers are getting their money's worth. Would the "smiles test" be sufficient?

I have tried to cover a lot of territory in this call to measurement and evaluation specialists to go public and do it well. I have described some of the current concerns associated with our profession and their legislative and legal ramifications. I have indicated that indeed we have no choice but to accept our public roles, and to pass on what we have learned to our juniors. At the same time, we cannot let the frustrations that will inevitably accompany our greater involvement in the public arena deflect us from continuously striving to improve our methods for assessing individuals and programs. I have contrasted the scientific codes for argument that we were taught to live by with the very different "rules" that govern the media and political groups. And I have suggested that our interactions with the public will be greatly enhanced if we learn to <u>recognize motivation</u>, <u>simplify communication</u>, and <u>minimize equivocation</u>.

We must also listen and attend to healthy criticism of measurement and suggestions for improvement that come from all quarters. However, many of the current attacks on testing do not fall in this category. Throughout, I have not attempted to downplay these out-and-out attacks. Now I would like to focus on them explicitly and label them for what they are: <u>vicious</u>, <u>destructive</u>, <u>deliberately misleading</u>, but also <u>sustained</u>,

well-organized, and well-financed. They play on the ignorance of the public and, if allowed to go unchecked, will undermine the whole foundation of our carefully developed and principled field f endeavor. On the surface, this may sound like a self-serving declaration. It is true that Educational Testing Service, as one of the largest measurement and research organizations in the field, has been the particular target of attacks. To quote ETS president, William Turnbull (1980, p. 2), "It is entirely proper that this organization, like any other, be subjected to public scrutiny. ETS must expect--and does expect--to be accountable for its own policies and performance." But it would be shortsighted indeed for you to think that any of you engaged in the profession, or the institutions you represent, will not be touched or affected by the attacks which, to quote Turnbull again, "rank among the most sustained and hostile assaults on an applied social science ever seen in this country." As individuals and through your professional associations, you cannot stand aside from the debate. Nor can you ignore your responsibilities to educate the public to the legitimate uses and social benefits of measurement and evaluation.

References

Jacobson, R. L. "Blacks lag in SAT scores." The Chronicle of Higher Education, January 7, 1980, p. 5.

Lerner, B. The war on testing: Detroit Edison in perspective. Princeton, N.J.: Educational Testing Service, 1979.

NAACP report on minority testing. N.Y.: NAACP Special Contribution Fund, 1976.

Turnbull, W. "APA and the war on testing." APA Monitor, March 1980, p. 2.

Williams, R. I., "Black pride, academic relevance and individual achievement." In R. W. Tyler and R. M. Wolf (Eds.), Crucial issues in testing. Berkeley: McCutchan Publishing Corporation, 1974.